Asian Journal of Applied Science and Technology (AJAST)
Volume 7, Issue 2, Pages 08-14, April-June 2023

Survey on Mortgage-Backed Securities Prepayment Risk Using Machine Learning Models

S. Parkavi[1*], Dr. P. Kanimozhi[2] & Dr. T. Ananth Kumar[3]

[1,2,3]Department of Computer Science & Engineering, IFET College of Engineering, Tamilnadu, India.
*Corresponding Author (S. Parkavi) - parkavi3152001@gmail.com**

DOI: https://doi.org/10.38177/ajast.2023.7202

Article Received: 17 February 2023 *Article Accepted: 28 March 2023* *Article Published: 14 April 2023*

ABSTRACT

Mortgages allow for partial or complete prepayment at any time prior to the loan's maturity date. A mortgage prepayment is risky for the lending institution because it eliminates the borrower's obligation to pay future interest and complicates the steps required for refinancing the mortgage loan. Prepayment risk is a threat to mortgage-backed securities (MBS) that arises when borrowers pay down their mortgage principal more quickly than expected. For MBS investors, this means lower cash flows. A negative impact on cash flows and exposure to extension risk could result from actual prepayment rates being higher or lower than expected.

Keywords: Mortgage; Prediction; Security; Banking.

1. Introduction

A mortgage loan has the opportunity to pay down a portion or the entire loan amount before it matures. This option is known as mortgage prepayment, and it puts the bank lending the mortgage loan at risk owing to the loss of future interest payments and the introduction of complexities in refinancing plans. Prepayment risk is a risk associated with mortgage-backed securities (MBS), in which borrowers pay off their mortgages sooner than projected, resulting in lower cash flows for MBS investors. Actual prepayment rates that are greater than anticipated can result in smaller cash flows, while actual prepayment rates that are lower than anticipated might expose investors to extension risk. The management of risk exposure and optimisation of investment strategies can be made easier for mortgage investors and lenders as a result of this research. Investments known as mortgage-backed securities (MBS) are those that are supported by a collection of mortgage loans. Prepayment risk, or the possibility that borrowers will repay their loans early than anticipated, is one of the main dangers connected to MBS. This may happen for several reasons, including refinancing, selling the house, or default.

In this survey, the prepayment risk of mortgage-backed securities will be predicted using machine learning. The model will include a variety of factors, including borrower demographics, loan characteristics, and macroeconomic indicators, and it will be trained on historical mortgage data [26]. The objective is to create a model and user interface that can predict prepayment rates with accuracy and pinpoint the major influences on prepayment behaviour. Using this data, pre-emptive tactics can be created to reduce the risk of prepayment, such as modifying loan terms or focusing marketing efforts on customers who are less likely to refinance [27]. In general, the creation of a machine learning model for anticipating prepayment risk in mortgage-backed securities has the potential to offer lenders and investors in MBS useful information and assist them in making wise investment choices.

2. Related Works

A credit rating model built on artificial neural networks was suggested by Byan-Jankar et al. [1] to identify possible charged-off loan applications. Prepayment and default behaviours for peer-to-peer loans are significantly

influenced by the length of the loan's age, the season, the rest phenomena, and the borrower's credit score. The authors G. Dong and K. K. Lai [2] suggested using random coefficients in logistic regression to increase prediction accuracy. Without losing desired characteristics, the suggested model can increase the predictive power of logistic regression. The suggested credit score card creation approach is anticipated to help manage credit risk effectively in real-world situations. A certain number of loans will result in either a profit for the bank or a loss for the bank depending on whether or not the borrower or customer pays back the loan. The banking industry is currently confronted with the most significant challenge, which is the recovery of loans. Using machine learning models that have been trained on the prior data set, the authors Singh, Vishal, et al. [3] make a prediction as to whether a new loan application will be approved or denied. In the course of their investigation into home loans with variable interest rates for Singapore's public housing. According to the findings of the investigation that Lee and Ong [4] carried out in the year 2003, prepayment demonstrated a positive link with market sentiment and the interest rate on loans, but a negative correlation with income growth and the relative price difference between private and public housing. According to the propositions that Xiong, Wei, and Yilong Han [5] present through two realistic cases that are consistent with the incomplete contract theory and the reference point theory, the findings show that the proposed compensation method increases the private sector's incentives for relationship-specific investments but deters its efforts to prevent early termination in serious risk scenarios. This is demonstrated by the fact that the proposed method increases the incentives for relationship-specific investments but decreases the private sector's efforts to prevent early termination. According to these findings, the proposed compensation method not only decreases the efforts made by the private sector to prevent early termination in serious risk scenarios, but it also increases the incentives for relationship-specific investments made by the private sector. This provides practitioners with answers for the early termination of public-private partnership projects, in addition to contributing to the theoretical advancement of public-private partnerships in contract design and incentive mechanisms.

In this Zhao, Hongke, et al. [6] reconstruct the two making of investment decisions phases of how to purchase and what amount to spend in order to concentrate on the personalised investment advice. To solve the "what to buy" challenge, the authors created for every investor, a list of potential investments. In this it considers several distinctive qualities of investment recommendations. The "how much money to pay" dilemma is also resolved by optimising and taking into account the investments a potential investor has made in each prospective candidate presently has, in accordance with portfolio theory. The approach is demonstrated by substantial experimental findings on a sizable real-world dataset using a variety of assessment measures.

The authors [7] developed criteria for identifying Chinese individual mortgage prepayment behaviour utilising the clustering method in order to solve the prepayment problem that was present in the securitization of individual mortgages. This was done in order to solve the problem of individual mortgages being securitized. In order to create a utilise the multi-period mortgage risk deep learning model to examine a historically large collection of monthly performance data and origination data. J.Sirignano, A.Sadhwani [8] data records for more than 120 million mortgages that started in the United States of America between early 90's to 20's available are taken. The unemployment has a significant impact on how sensitive a borrower is to fluctuations in unemployment. Additionally, it greatly varies among the overall borrower population, highlighting how unemployment interacts

Asian Journal of Applied Science and Technology (AJAST)
Volume 7, Issue 2, Pages 08-14, April-June 2023

with several other factors. The authors [9] outline the development of creating a machine learning prediction model which can benefit banks identify borrowers who are qualified using financial information to submit a loan application. By using machine learning models and artificial intelligence, the recommended model is made so that bankers can only extract entries that have been electronically signed for, have a range of average credit ratings, and have a limited amount of loan payback.

The Studies from outside are relatively advanced in relation to of the elements that affect the early mortgage loan payback. Daniel (2010) examined the home loan market in Australia and believed that the as the loan pool's age increased a significant influence on full prepayment [10].

Daniel [11], in his research on the housing loan market in Australia, came to the conclusion that the average age of the loan pool had a significant influence on prepayment in full. The goal of sparse classification algorithms is to learn a classifier that is as sparse as is practically possible. Frequently, some prior assumption about the weights that encourages their sparsity regularises the probability of the weights when training data is present. When performing logistic regression with the automated relevance determination paradigm, the prior may at times be implicit. The authors Yue, and Liu Lan [12] begins with the concept of marketizing interest rates, examines impact of rate of interest marketization on merchant banks' interest rate endeavour from both the on- and off-balance sheet perspectives of merchant banks, and suggests remedies. As the foundation of the financial system, merchant banks are well recognised to rely on net interest income The volatility and uncertainty of interest rate levels have gotten worse as interest rate marketization in China continues to advance. The stability of China's financial market will be impacted by interest rate venture, which had become a significant risk for China's merchant banks.

Ashwitha, K., et al. [13] Created a machine learning loan prediction system therefore, which will choose the most appropriate applicants when they apply for a loan and will also determine their eligibility. This will be advantageous to the candidate as well considering the bank staff. The timeline for authorising loans will be significantly shortened. By offering a variety of financial solutions to their clients, financial institutions are concentrating on increasing their revenue streams, a significant portion of this income comes from the business of credit lines. Darapaneni, Narayana, et al [14] sorts out the profitability of a financial institution depends on how successfully the credit company generates income, there is a strong emphasis on optimising this process and a strong desire to lower the risk of loan defaulters. An appropriate hierarchical GUI structure is generated using a data-driven K-nearest-neighbors technique, from which a prototype application shall become immediately put together. Through this strategy was put into practise for Android via a programme named ReDraw. According to Moran, Kevin [15] study, redraw correctly categorises graphical user interface components using an average accuracy and develops prototype mobile applications that, in terms of visual affinity, fairly approximate target mock-ups. Redraw may improve actual development processes, according to interviews with industry users. Most of the time, to sanction legal approval, bank managers and lawyers must all bulk bundles of registration paperwork should be checked from prior years. To complete the entire procedure, lawyers grasp their time to draw conclusions by reviewing every document. The authors Rajesh, P., et al. [16] tried to reduce the risk factor, computational time, and cross-verification by means of data science decision tree approach. Abedin, Mohammad Zoynul [17] suggests employing an automated approach for tax default prediction to deal with the financial effects of unpaid taxes. It

mainly emphasises the critical need of properly developed systems that forecast tax defaults can call for a mix of strategies for machine learning and transformation of data methods. The use of efficient automated system for predicting tax defaults has significant consequences for tax administration and can help managers reach realistic allocations of government spending and revenue growth.

The necessity to arrange textual documents and the exponential development in their production have raised awareness of the automatic categorization of texts into predetermined categories. There are several supervised learning techniques that deal with text classification. Moldagulova Aiman, and Rosnafisah Bte Sulaiman [18] described a way for developing a R-based machine learning system employs the K-Nearest Neighbours' method to the categorization of textual documents. The K-Nearest neighbour's algorithm has been devoted to the tough task of determining the suitable value for the indicator of the number of neighbour's k. To the restricted number of individuals who qualify, the bank will often offer based on net property loans worth also bank assertions, with loan approvals depending on the customers' income resources. Kumar, Ch Naveen [19] determined if a consumer will be approved for a loan based on techniques of machine learning. In order to analyse the data pertaining to customers the factors that are necessary to combine making use of machine learning methods, client data is gathered based on multiple banks and accessed through customer profiles with KNN algorithm.

Due to the multicollinearity of the interpretive variables in regression models, social post-evaluation is typically challenging. Chen, Li [20] suggested the benefits of ridge regression over the LS method. The most cutting-edge machine learning method for data mining is the support vector machine, which forms the foundation for structural risk mitigation and has demonstrated to be more compared to the more prevalent empirical risk reduction. The algorithms such as support vector machines and ridge regression were used for this work to analyse the world bank projects. The combination's efficacy is supported by experimental results and theoretical studies. The issue of developing algorithms that make predictions to determine the cost of residences that are sold in large cities yet challenging and demanding. For those places, Numerous interconnected factors affect the sale price of real estate. The price may be influenced by the property's features, location, and size, among other factors. By displaying the accessible housing properties on a machine hackathon platform, Manasa, J., Radha Gupta [21] had analytical research conducted by considering the data set which are still available to the general public. The goal is to develop a prediction model for cost evaluation based on characteristics that influence pricing. Lin, Weiwei, [22] use a bootstrap sampling method with heuristics along with the ensemble learning algorithm on the large-scale insurance business data mining, and suggested a random forest ensemble technique that makes use of Spark's memory-cache optimisation and parallel computing capability. To analyse potential clients using the suggested technique, obtained the sector of insurance the China Life Insurance Company's statistics. Moreover, it is advantageous for enhancing product marketing accuracy as compared to the conventional artificial strategy.

As a characteristic of the real estate market, fluctuations are certain to have an effect on the financial system, posing significant dangers to the economy and potentially resulting in a financial crisis if not properly managed or controlled. Based on an examination of mortgage risks associated with home purchases and loan risks associated with real estate development, Yang, Xiaozhuang [23] investigates the origins and drivers of financial hazards. It also offers policy recommendations for mitigating these risks. Typically occurring problem in the modern at this

time, credit card theft is being detected worldwide. The issue is brought on by an increase in e-commerce platforms and online transactions. When a credit card is used for any unauthorised activity, even when the fraudster utilises the card's details for his own gain, credit card fraud often results. Sailusha, Ruttala [24] analyses the credit card fraud detection system allowed for the detection of fraudulent actions. The main objective of this study is to concentrate on using machine learning techniques. Both the Adaboost method and the random forest algorithm are employed.

3. Problem Statement

Mortgage-Backed Securities (MBS) are investment vehicles that are backed by pools of mortgages. The prepayment risk associated with MBS is a significant concern for investors, as it can result in reduced returns or losses. This project's goal is to create a machine learning-based system that can predict prepayment risk for MBS and help investors make more informed decisions.

4. Conclusion

In summary, Mortgage-Backed Securities (MBS) are financial products made by combining mortgages and selling them to investors. MBS have grown in popularity over time as a result of their high yield and relative safety when compared to other investments. Prepayment risk exists when investing in MBS, nevertheless. Prepayment risk refers to the possibility that borrowers will settle their debts ahead of schedule, which could result in losses for investors. Due to the increased likelihood of mortgage refinancing during periods of low interest rates, this risk is particularly substantial. Overall, this effort demonstrates the potential of machine learning to enhance MBS prepayment risk modelling and deliver more precise and trustworthy predictions. Accurate prepayment risk assessment can result in improved investment decisions and risk management, which has significant consequences for investors and financial institutions involved in the MBS market. We are aware of the limitations of our models, though, and future work can look into ways to increase their precision and resilience.

Declarations

Source of Funding

This study did not receive any grant from funding agencies in the public or not-for-profit sectors.

Competing Interests Statement

Authors have declared no competing interests.

Consent for Publication

The authors declare that they consented to the publication of this study.

References

[1] Byanjankar, M. Heikkilä, and J. Mezei (2015). Predicting credit risk in peer-to-peer lending: A neural network approach. In Proc. IEEE SSCI, Pages 719–725.

[2] G. Dong, K.K. Lai, and J. Yen (2010). Credit score card based on logistic regression with random coefficients. Procedia Comput. Sci., 1(1): 2463–2468.

[3] Singh, Vishal, et al. (2021). Prediction of modernized loan approval system based on machine learning approach. International Conference on Intelligent Technologies.

[4] Luo, H. Xiong, W. Zhou, Y. Guo, and G. Deng (2011). Enhancing investment decisions in P2P lending: An investor composition perspective. In Proc. ACM SIGKDD, Pages 292–300.

[5] Xiong, Wei, and Yilong Han (2021). Incentives of Early Termination Compensation in Public–Private Partnership Projects. IEEE Transactions on Engineering Management.

[6] Zhao, Hongke, et al. (2014). Investment recommendation in p2p lending: A portfolio perspective with risk management. IEEE international conference on Data Mining.

[7] Pavithra, M., R. Rajmohan, T. Ananth Kumar, and R. Ramya (2021). Prediction and classification of breast cancer using discriminative learning models and techniques. Machine Vision Inspection Systems, Volume 2, Machine Learning-Based Approaches, Pages 241–262.

[8] Sirignano, Justin, Apaar Sadhwani, and Kay Giesecke (2016). Deep learning for mortgage risk. ArXiv preprint arXiv:1607.02470.

[9] Ragapriya, N., T. Ananth Kumar, R. Parthiban, P. Divya, S. Jayalakshmi, and D. Raghu Raman (2023). Machine Learning Based House Price Prediction Using Modified Extreme Boosting. Asian Journal of Applied Science and Technology, 7(1): 41–54.

[10] Daniel J.A. (2010). Fixed-rate loan prepayment model for Australian mortgages [J]. Australian Journal of Management, 35(1): 99–112.

[11] Krishnapuram, Balaji, et al. (2005). Sparse multinomial logistic regression: Fast algorithms and generalization bounds. IEEE transactions on pattern analysis and machine intelligence, 27(6): 957–968.

[12] Yu, Yue, and Liu Lan (2019). The Impact of Interest Rate Marketization on the Interest Rate Risk of Commercial Banks. International Conference on Data Science and Business Analytics.

[13] Ashwitha, K., et al. (2022). An Approach to Predict Loan Eligibility using Machine Learning. International Conference on Artificial Intelligence and Data Engineering.

[14] Darapaneni, Narayana, et al. (2022). Loan prediction software for financial institutions. Interdisciplinary Research in Technology and Management.

[15] Moran, Kevin, et al. (2018). Machine learning-based prototyping of graphical user interfaces for mobile apps. IEEE Transactions on Software Engineering, 46(2): 196–221.

[16] Rajesh, P., et al. (2020). Real Time Data Science Decision Tree Approach to Approve Bank Loan from Lawyer’s Perspective. IEEE International Conference on Machine Learning and Applications.

[17] Abedin, Mohammad Zoynul, et al. (2020). Tax default prediction using feature transformation-based machine learning. IEEE Access, 9: 19864–19881.

[18] Moldagulova, Aiman, and Rosnafisah Bte Sulaiman (2017). Using KNN algorithm for classification of textual documents. International Conference on Information Technology.

Asian Journal of Applied Science and Technology (AJAST)
Volume 7, Issue 2, Pages 08-14, April-June 2023

[19] Kumar, Ch Naveen, et al. (2022). Customer Loan Eligibility Prediction using Machine Learning Algorithms in Banking Sector. International Conference on Communication and Electronics Systems.

[20] Chen, Li (2011). Social Post-Evaluation of World Bank Projects in Yanhe Basin Based on Ridge Regression and Support Vector Machines. International Workshop on Intelligent Systems and Applications.

[21] Manasa, J., Radha Gupta, and N.S. Narahari (2020). Machine learning based predicting house prices using regression techniques. International Conference on Innovative Mechanisms for Industry Applications.

[22] Lin, Weiwei, et al. (2017). An ensemble random forest algorithm for insurance big data analysis. IEEE Access, 5: 16568–16575.

[23] Yang, Xiaozhuang (2008). Real estate financial risk prevention and countermeasure. International Conference on Wireless Communications, Networking and Mobile Computing.

[24] Sailusha, Ruttala, et al. (2020). Credit card fraud detection using machine learning. International Conference on Intelligent Computing and Control Systems.

[25] Zhiying, Wang, and Sun Jing (2006). Application of DMAIC on service improvement of bank counter. International Conference on Service Systems and Service Management.

[26] Jabez, L.K., K. Suresh Kumar, G. Ganesh, and M. Ajai (2021). Design and Implementation of Artificial Intelligence Based Power Management System for industrial Application. International Conference on System, Computation, Automation and Networking, Pages 1–6.

[27] Adeniyi, Jide Kehinde, Abidemi Emmanuel Adeniyi, Yetunde Josephine Oguns, Gabriel Olumide Egbedokun, Kehinde Douglas Ajagbe, Princewill Chima Obuzor, and Sunday Adeola Ajagbe (2022). Comparison of the Performance of Machine Learning Techniques in the Prediction of Employee. ParadigmPlus, 3(3): 1–15.